

Discussion paper / Artículo de reflexión / Documento de discussão - Tipo 2

Towards Smarter Cities Taking Advantage of the Fog Computing Paradigm

Shouddy Tárano León / shouddy@gmail.com

Tatiana Delgado Fernández / tdelgado@ind.cujae.edu.cu

Alejandro Luar Pérez Colomé / alejandroluar@gmail.com

Universidad Tecnológica de La Habana, CUJAE - Cuba

ABSTRACT The fog computing term has achieved importance in the last years due to its effect in the latency reduction that the Internet of Things [IoT] applications have. These applications demand real-time (or nearly real-time) responses and they are characterized by low bandwidth consumption; hence, the fog computing is relevant in achieving these requests because part of the processing is done near the end user devices. For this reason, the cloud computing paradigm is not enough for some applications, since nowadays, the instant need of data and the decision-making process leverage –or somehow discover– a new horizon that demands a complementary variable. This article consists on an approach to the fog computing term, together with the requirements analysis for engineering solutions in the IoT field. Also, its impact in the smart cities and other fields plus its main challenges are addressed. We also present a guideline to implement a recommendation system for sightseeing places for tourists based in fog computing embraced in a large smart city project located in Havana.

KEYWORDS Fog computing; cloud computing; smartcities.

Hacia ciudades más inteligentes aprovechando el paradigma de Fog Computing

RESUMEN El término fog computing [computación en la niebla o niebla computacional] ha cobrado auge en los últimos años por su incidencia en la disminución de la latencia que tienen las aplicaciones de Internet de las Cosas [IoT, Internet of Things], las cuales demandan respuestas en tiempo real o cercano al real, así como por el menor consumo de ancho de banda que resulta de resolver parte del procesamiento más cerca de los dispositivos del usuario final. Ya no es suficiente el paradigma de la cloud computing [nube computacional]. En el presente, la necesidad del “dato” y de la toma de decisión al instante impulsan, o de alguna manera descubren, un horizonte nuevo que demanda de una variante complementaria. Este artículo constituye un acercamiento al término fog computing, aparejado con el análisis de su necesidad ante soluciones ingenieriles en el campo de la IoT, su impacto en las ciudades inteligentes y demás campos de acción, y algunos de sus principales retos. Se ofrece además, una hoja de ruta para implementar un sistema de recomendación de lugares de interés al viajero basado en fog computing, en el marco de un proyecto de ciudades inteligentes en La Habana.

PALABRAS CLAVE Niebla computacional; computación en la nube; ciudades inteligentes.

Rumo a cidades mais inteligentes aproveitando o paradigma Fog Computing

RESUMO O termo fog computing [Computação em neblina ou neblina computacional] ganhou impulso nos últimos anos devido à sua incidência na diminuição da latência que as aplicações da Internet das Coisas [IoT, Internet of Things] têm, as quais demandam respostas em tempo real ou próximo do real, bem como o menor consumo de largura de banda que resulta da resolução de parte do processamento mais próximo dos dispositivos do usuário final. O paradigma da computação em nuvem já não é suficiente. No presente, a necessidade do “dato” e de tomada de decisão instantânea, conduzem, ou de alguma forma descobrem, um novo horizonte que exige uma variante complementar. Este artigo é uma abordagem do conceito de fog computing, juntamente com a análise de sua necessidade de soluções de engenharia no campo da IoT, seu impacto em cidades inteligentes e outros campos de ação e alguns dos seus principais desafios. Ele também oferece um roteiro para implementar um sistema de recomendação de locais de interesse para o viajante com base na fog computing, no âmbito de um projeto de cidades inteligentes em Havana.

PALAVRAS-CHAVE Névoa computacional; computação em nuvem; cidades inteligentes.

I. Introduction

Several periods have been written into the introduction of computing to the society. Although the authors tend to differ between the exact temporary bounds, there is a general consensus in the main stages in the connectivity models and services when technological challenges have arisen. Three stages have been identified:

- The first one, in the 60s until the early 80s, when the mainframe and the dumb terminals with green screens allowed to execute the batch processing in banks, universities and others;
- the second stage, in the 80s, when a new client-server stage started. This one –together with the revolution and popularization of the web– allowed the duration of this stage until the end of the 20th century and it entailed new technological developments relative to the computing technologies and data storage; and
- the third one, thanks to the popularization of mobile devices and cloud computing. This occasioned the industrialization of the computing and data storage businesses, such as the development environments and consumption of end-services from the web –e.g., the email– (De Fuenmayor,2017).

The reader might notice that, at the end-user level and its respective needs, an adequate balance between the services demand that the internet provides is present. Even though the internet is not a service with world-wide coverage, there are several application fields from pure science to the solely entertainment, which includes large amounts of people in countries where there is a generalized access to the service.

The bandwidth needs in many countries are covered by the average user (see **FIGURE 1**). As per a study performed by Akamai Technologies (2017), the speeds per subscriber are up to 7.2 Mbps, with an annual growing of 15% in the first quarter of 2017, compared with the same time in 2016 (De Fuenmayor, 2017).

This leads us to the following questions:

- Does the cloud computing comply with all the emerging solution needs, not only the connectivity demands of the end-user devices (smartphones, tablets, PCs), but also devices that can transmit information of their automation objects in an autonomous way?
- Which are the alternatives to the information processing need emitted by these devices for an opportune, effective, and efficient decision-making process?

I. Introducción

En el proceso de la introducción de la computación en la sociedad se han descrito varios períodos. Aunque los autores tienden a diferir entre las fronteras temporales exactas, existe consenso en las etapas generales cuando se han presentado los retos a nivel tecnológico, en cuanto a modelos de conectividad y servicios. Se identifican tres etapas:

- la primera, entre los años sesenta y ochenta, cuando el *mainframe* y los terminales tontos con pantalla verde permitían ejecutar el procesamiento “por lotes” en bancos, universidades, etcétera;
- la segunda, en los años ochenta, cuando comenzó una nueva etapa cliente-servidor que, junto con la revolución y popularización de la Web, consiguió extenderse hasta finales del siglo XX y marcó un nuevo desarrollo tecnológico en torno a las tecnologías de la computación y el almacenamiento de datos; y
- la tercera, que llegó gracias a la popularización de los dispositivos móviles y al cloud computing, lo que supuso la industrialización del negocio de la computación y el almacenamiento de datos, así como los entornos de desarrollo y consumo de servicios finales desde la red, como el correo electrónico (De Fuenmayor,2017).

Hasta este momento se puede notar que, a nivel de usuario final y de sus necesidades, se presenta un balance adecuado entre la demanda y los servicios que la Internet brinda. Aunque la Internet dista de ser un servicio con una cobertura a todos los rincones del planeta, existen diferentes campos de aplicación, desde la ciencia pura, hasta el mero entretenimiento, que incluyen a grandes masas en los países donde hay acceso generalizado al servicio.

Las necesidades de ancho de banda en muchos países están cubiertas para el usuario medio (ver **FIGURA 1**). Según un estudio realizado por Akamai Technologies (2017), las velocidades a nivel de usuario global suscrito, llegan hasta los 7.2 Mbps, con un crecimiento anual de 15 % en el primer cuatrimestre del 2017, comparado con el mismo momento del 2016 (De Fuenmayor, 2017).

Lo expuesto lleva a las siguientes interrogantes:

- ¿Cumple el cloud computing todas las necesidades emergentes de solución, no solo ante la demanda de

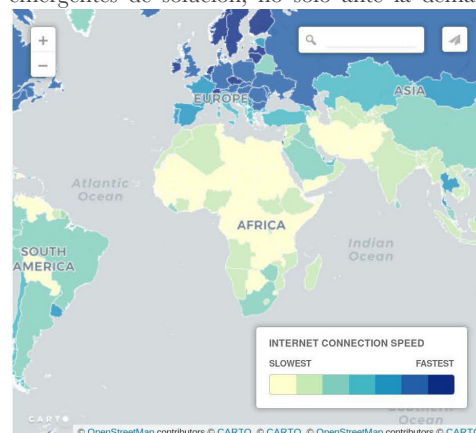


Figure 1. Global internet speed / Velocidad de Internet a nivel global (“Mapped...”, 2017)

conectividad de las estaciones finales de usuario (*smart phones*, tabletas, PC), sino de dispositivos que pueden emitir información de sus objetos de automatización de manera autónoma?

- ¿Cuáles son las alternativas que se presentan ante la necesidad de procesamiento de la información emitida por estos dispositivos, para una toma de decisión oportuna, efectiva y eficiente?

II. El nuevo actor: la Internet de las Cosas

Cisco Systems (2017) acuña la IoT [*Internet of Things*] como la red de objetos físicos o “cosas” electrónicas, formada por programas, sensores y conectividad, que se utilizan para generar valor y servicio mediante el intercambio de datos con el fabricante, el operador u otros dispositivos conectados a través de protocolos de comunicación avanzados sin operación humana (Abdelshkour, 2015). Este entorno se ha desarrollado aparejado con el desarrollo de las TIC y la infraestructura tecnológica de las sociedades.

La IoT brinda acceso instantáneo a un dispositivo, asegurando su adecuada instalación y comportamiento, el análisis de los datos obtenidos, la conexión interoperable con su LAN [*Local Area Network*], nube u otros dispositivos conectados por cualquier medio en una red inalámbrica o cableada, según sea técnicamente adecuado (Bojanova, 2015).

Hasta ahora, el modelo seguido para la adquisición y el análisis de la información recogida del IoT se basa en el paradigma de computación en la nube, sin embargo, este modelo no deja de ser una instancia de una arquitectura centralizada tradicional, con grandes (y escasos) centros de datos y con conectividad de primer nivel. Este modelo no es escalable ni adecuado para los retos que el IoT pone sobre la mesa a corto-medio plazo, como asevera Arco Research Group (2017): “La transmisión de la cantidad de datos generados por los dispositivos IoT a una única localización para ser procesados no será técnica ni económicamente viable”.

Aunque las ventajas de la nube computacional soportan e impulsan el desarrollo de la interacción entre las Tecnologías de la Información y las Comunicaciones [TIC] y la sociedad en todas sus dimensiones, es cierto que ante la cantidad de equipos autónomos con lógica computacional de mayor o menor complejidad, ha irrumpido necesariamente en la vida de los ciudadanos y las organizaciones.

El enfoque, desde el punto de vista de la ingeniería, se ha encontrado con algunas limitaciones ante el creciente aumento de la demanda de conectividad, entre ellas la latencia y el ancho de banda.

- Latencia: la distancia de red entre los dispositivos y los centros de procesamiento está en torno a centenares de milisegundos (ida y vuelta). Muchas de las aplicaciones en el IoT serán sensibles a la latencia de la red, como por ejemplo, los sistemas de predicción (consumo energético y las *smart grids*) y los sistemas de control por voz (Apple Siri, Google Home o Alexa de Amazon).
- Ancho de banda: se prevé una contribución cada vez mayor de información proveniente de dispositivos como

II. The New Actor: Internet of Things

Cisco Systems (2017) coins the IoT [*Internet of Things*] as the network of objects –i.e., electronic “stuff”– formed by programs, sensors, and connectivity used to generate value and service towards data exchange with the manufacturer, the operator, or another connected devices through advanced communication protocols without human intervention (Abdelshkour, 2015). This environment has been developed by pairing the ICT [Information and Communication Technologies] development and the technological infrastructure of the society.

IoT grants instant access to a device by guaranteeing its correct installation and behavior, the analysis of the obtained data, the interoperable connection with its corresponding LAN [*Local Area Network*], cloud, or other connected devices by any medium in a wired or wireless network as it is technically suitable (Bojanova, 2015).

Until now, the followed model for the gathering and analysis of the information picked from IoT is based on the cloud computing paradigm. Nonetheless, this model is no more than an instance of a traditional and centralized architecture with large (and limited) datacenters with first-level connectivity. This model is not scalable nor suitable for the challenges the IoT entails in the short and medium term, as the Arco Research Group (2017) indicates: “the transmission of the amount of data generated by IoT devices to a single location to be processed will not be technically nor economically viable”.

Despite the advantages of the cloud computing supporting and leveraging the development of the interaction between the ICT and the society in all its aspects, it is true that the number of autonomous devices with computational logic –either with low or high complexity– has invaded the life of the citizens and the organizations.

The approach from the engineering point of view has found some limitations to the growing in the connectivity demand, such as the latency and bandwidth.

- Latency: the network distance between the device and the processing central is about hundreds of milliseconds (round-trip). Many of the IoT applications will be latency-sensitive such as the predicting systems (energy consumption and smart grids) and the voice controlled systems (Siri, Google Home, or Alexa).
- Bandwidth: a larger information contribution coming from devices processing audio and video is predicted, especially in cities. Surveillance cameras

in public/private locations, traffic monitoring, and personal phones are some examples. The information volume to process in real-time and send it to the cloud would result in inadmissible response times, if the technological capacity required for that is on its place (Abdelshkour, 2015).

One of the options to face these realities is the so-called fog computing (Abdelshkour, 2015). The term –also coined by Cisco– establishes a “fog” model formed by a distributed collection and data analysis points generated by ads, entertainment, computing, and other aspects featuring the informatic applications housed in any network-enabled device (Bonomi, Milito, Zhu, & Addepalli, 2012).

Within the fog computing model, the available computing resources can be used in the end devices or in nodes not being employed for general purposes. Alternatively, some computing resources can be added to the existing nodes or to an edged network to ease the processing near the end user device.

Fog computing has a group of associated advantages. The first one is the reduction in the network traffic by providing a platform for the filtering and analysis of the data generated by the sensors. This is done by using resources of the devices allocated in the edge, which drastically reduces the traffic sent to the cloud. Another advantage is the reduction in the latency, especially for applications requiring real-time processing.

Hu, Dhelim, Ning, & Qiu (2017) present another advantages of the fog relative to the location sensitiveness, the geographic distribution, the low energy consumption, and the security and protection of the data privacy. The fog computing supports the mobility demands based on the location and enables the administrators the power to control users’ location and the information access from mobile devices.

The decentralized architecture of the fog computing ensures the proximity to the client in the data analysis, which allows to perform faster big data analysis, to deploy better services based on the location, and to execute more powerful capacities for the real-time decision-making processes. Due to the deployment of services near the end user, this model presents considerable advantages in relation to the data security and the protection of the user privacy.

As the main disadvantage, it has been informed that this model might have a reduced tuning fork of technological platforms for development and construction,

frente de imagen o vídeo, especialmente en la ciudad. Las cámaras de seguridad en espacios públicos y privados, la monitorización del tráfico y los teléfonos personales son algunos de los ejemplos. El volumen de información a mover en tiempo real hacia la nube daría lugar a tiempos de respuesta inadmisibles, asumiendo que existiera la capacidad tecnológica para ello (Abdelshkour, 2015).

Una de las opciones que se ha encontrado para enfrentar estas realidades es el *fog computing* (Abdelshkour, 2015) el término, acuñado por Cisco (Bonomi, Milito, Zhu, & Addepalli, 2012), establece un modelo de “niebla” que se caracteriza por formarse a partir de puntos distribuidos de colección y análisis de datos, generados por anuncios, entretenimiento, cómputo y otros aspectos que caractericen las aplicaciones informáticas alojadas en cualquier dispositivo conectado a la red.

En el modelo de *fog computing* se pueden usar recursos de computación disponibles en los dispositivos de usuarios o en nodos que actualmente no se están usando para propósitos generales. Alternativamente, pueden ser añadidos recursos computacionales adicionales a los nodos existentes o a una red salto desde el borde, para facilitar el procesamiento más cerca del dispositivo del usuario (Varguese et al, 2017).

Fog computing tiene un grupo de ventajas asociadas (Gupta, Vahid-Dastjerdi, Ghosh, & Buyya, 2017). La primera es la reducción del tráfico en la red, al brindar una plataforma para el filtrado y análisis de los datos generados por los sensores, utilizando recursos de los dispositivos que están en el borde (*edge*), lo que reduce drásticamente el tráfico enviado a la Nube. Otra ventaja es la reducción de la latencia, especialmente para aplicaciones que requieren procesamiento en tiempo real.

Hu, Dhelim, Ning, y Qiu (2017) ofrecen otras ventajas de la niebla en relación con la sensibilidad a la ubicación, la distribución geográfica, el bajo consumo de energía y la seguridad y protección de la privacidad. La computación en la niebla soporta las demandas de la movilidad con base en la ubicación y habilita a los administradores para controlar dónde se encuentran los usuarios y los dispositivos móviles y cómo acceden a la información.

La arquitectura descentralizada de *fog computing* asegura la proximidad de la analítica de datos al cliente, lo que permite realizar análisis de *big data* más rápidos, desplegar mejores servicios basados en la ubicación y ejecutar capacidades más poderosas para la toma de decisiones en tiempo real. Debido al despliegue de los servicios cerca del usuario final, este modelo es también ventajoso en relación con la seguridad del dato y con la protección de la privacidad del usuario.

Como desventaja fundamental se ha descrito que este modelo podría contar con un reducido diáspora de plataformas tecnológicas de desarrollo y construcción. Así como su presentación a la hora de establecer estándares de interoperabilidad para alcanzar mayorías dentro de la población de dispositivos (Abdelshkour, 2015).

Este punto realmente impacta contra el avance que ha tenido la estandarización de soluciones y plataformas de desarrollo dentro del entorno de la nube computacional.

Table 1. Fog vs. cloud / Niebla vs nube (Abdelshkour,2015)

Requirements	Fog computing	Cloud computing
Latency	High	Low
<i>Latencia</i>	<i>Alta</i>	<i>Baja</i>
Delay in the jitter	High	Very low
<i>Demora en los jitters</i>	<i>Alta</i>	<i>Muy baja</i>
Service location	Inside internet	At the edge of local networks
<i>Ubicación del servicio</i>	<i>Dentro de Internet</i>	<i>Al borde de las redes locales</i>
Client-server distance	Multiple hops	Single hop
<i>Distancia cliente-servidor</i>	<i>Múltiples saltos</i>	<i>Salto único</i>
Security	Undefined	Defined
<i>Seguridad</i>	<i>Indefinida</i>	<i>Definida</i>
Attack in routed data	High probability	Very low probability
<i>Ataque en datos enrutados</i>	<i>Alta probabilidad</i>	<i>Muy baja probabilidad</i>
Location dependent	No	Yes
<i>Dependencia de la localización</i>	<i>No</i>	<i>Si</i>
Geo-distribution	Centralized	Distributed
<i>Geodistribución</i>	<i>Centralizada</i>	<i>Distribuida</i>
Number of servers	Low	Many
<i>Número de servidores</i>	<i>Bajo</i>	<i>Muchos</i>
Mobility support	Limited	Supported
<i>Soporte a la movilidad</i>	<i>Limitada</i>	<i>Soportada</i>
Real-time interaction	Supported	Supported
<i>Interacción en tiempo real</i>	<i>Soportada</i>	<i>Soportada</i>
Last-mile connectivity type	Dedicated line	Wireless
<i>Tipo de conectividad última milla</i>	<i>Línea dedicada</i>	<i>Inalámbrica</i>

Table 2. Features of the fog solutions compared to the cloud ones, expected results and behaviors / Características de las soluciones de niebla frente a las de nube, resultados y comportamientos esperados (Abdelshkour, 2015)

Cloud computing	Fog computing
The data and the applications are processed in the cloud. For large information volumes, its processing can be delayed.	Fog operates at the edge of devices. In this layer, less processing time is required.
<i>Los datos y las aplicaciones se trabajan en la nube, para grandes volúmenes de información su procesamiento puede ser demorado.</i>	<i>Fog opera en el borde de la red de los dispositivos, en esta capa, se consume menos tiempo de procesamiento.</i>
The employed cloud channel bandwidth can be compromised.	Less bandwidth demand, since data are aggregated in some intermediate device before being transported to the cloud, if this latter is required.
<i>El ancho de banda usado para el canal de la nube puede verse comprometido.</i>	<i>Menos demanda de ancho de banda pues los datos son agregados en algún dispositivo intermedio antes de ser transportado a la nube, de ser preciso hacerlo.</i>
Delays in the response time and scalability problems due to the servers' location in the deployment of the solution.	it is possible to encourage less delays in the final response times and less scalability issues in the deployment through border servers determined by the users' visibility in the network.
<i>Demoras en el tiempo de respuesta y problemas de escalabilidad debido a la localización de los servidores en el despliegue de la solución.</i>	<i>Por medio de servidores de borde determinados por la visibilidad de los usuarios en la red, se pueden fomentar menos demoras en los tiempos de respuesta finales y menores problemas de escalabilidad del despliegue</i>

such as its presentation when the interoperability standards are established to reach the majority of devices (Abdelshkour, 2015). This topic really impacts the advance that the standardization of solutions and platforms have had within the computing cloud. Cisco shows a summary of the aspects to be considered when fog and cloud are put in a weighting scale (see **TABLE 1**). On the other hand, the role of the servers and the information processing are shown in **TABLE 2**.

III. Architecture and Quality of Service in Fog Computing

Understand on how to improve the QoS [Quality of Service] in the IoT networks is becoming a challenge. The fog computing approach is based on getting the end users as close as possible to the cloud to improve the general performance without being limited to a particular architecture (such as the cellular networks). The fog computing is an intelligent layer located between the cloud and the IoT, which provides low latency, location knowledge, and generalized geographic distribution for the IoT (see **FIGURE 2**).

The basic way the IoT, fog, and cloud nodes operate and interact between them is exposed in the next paragraphs. The IoT nodes process their requests locally, sending it after to a fog or cloud node. After that, the fog nodes can process the received requests; if they cannot, they forward the task to another fog node in the same domain. If these latter are unable to process the requests, it is sent to the cloud, where the nodes there will process the requests and send the responses to the IoT nodes (Yousefpour et al., 2017).

All the steps entail delays in the service, which is no more than the required time to attend a request; that is, the time interval between the time a IoT node sends a service request and the moment when it receives response from that request.

The decision to perform a task is based on the response time of a fog node, which depends of several factors: the amount of

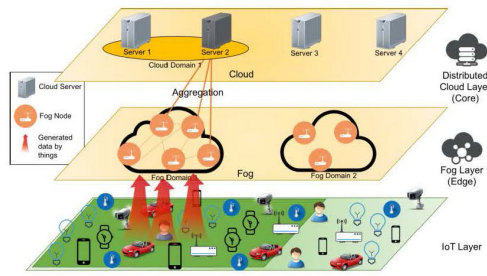


Figure 2. General frame for the IoT-fog-cloud architecture / Marco general para la arquitectura IoT-niebla-nube (Yousefpour, Ishigaki, & Jue, 2017)

computing power required for the task, the state of the queue, and the processing capacity of a fog node. It is important to consider the different processing times according to the various tasks. In other words, the heavy and light processing tasks should be distinguished (Chin-Feng, Dong-Yu, & Ren-Hung, 2016). For instance, the requests sent by temperature sensors towards fog nodes to calculate the average room temperature can be seen as light processing tasks; whilst a request sent from a video camera to read the license plate of a vehicle to the fog nodes is an examples of a heavy processing task.

When the fog nodes receive requests from IoT nodes participating in an application, they must distinguish between light and heavy requests. For that, the requests must have a field identifying the request type on their header; e.g., the traffic class field in the IPv6 header.

There are two interaction modes for fog nodes: a centralized mode, where a central authority controls the interaction between the fog nodes; and a distributed one, where the fog nodes interact with their neighbors using a universal protocol (Cardellini, Grassi, Lo Presti, & Nardelli, 2015).

The centralized interaction mode can be seen as a central resource orchestration, where the central node –which knows the topology and the state of the fog nodes in a domain– is present. The centralized mode is easier to implement, since there is no need of a distributed communication through certain protocols. Furthermore, the central node might be used as a medium to push the fog applications and perform updates to the nodes within its domain.

The distributed mode is the more suitable for scenarios where the fog nodes are not static, or when the fog network is designed in an ad hoc manner. Besides, in this mode, there is no need of having a dedicated node to

Cisco muestra un resumen de los aspectos a tener en cuenta a la hora de poner en una balanza la niebla y la nube (*fog vs cloud*) computacionales (ver Tabla 1). Por otra parte, el rol de los servidores y el procesamiento de la información, en ambos conceptos, se muestra en la Tabla 2.

III. Arquitectura y calidad del servicio en *fog computing*

Entender cómo mejorar la calidad de servicio [*Quality of Service*, QoS] en las redes IoT se está transformando en un desafío. La computación de niebla apunta a acercar la nube a los usuarios finales para mejorarla, sin estar limitado a una arquitectura en particular (como la red celular). La computación de niebla es una capa inteligente, asentada entre la nube y la IoT, que trae baja latencia y el conocimiento de la ubicación y distribución geográfica generalizada para el IoT (ver FIGURA 2).

La forma básica en que los nodos de IoT, los nodos de niebla y los nodos de la nube operan e interactúan entre sí se expone a continuación: los nodos de IoT procesan su solicitud localmente, enviándola luego a un nodo niebla o a un nodo de la nube. Posteriormente, los nodos niebla pueden procesar las solicitudes recibidas, si le es imposible, en ese instante, realizar esa tarea, reenvían la solicitud a otro nodo niebla del mismo dominio. Si estos tampoco pueden procesarla, se reenvía la solicitud a la nube, donde los nodos en la nube procesan las solicitudes y envían la respuesta a los nodos IoT (Yousefpour et al., 2017).

Todos los pasos planteados incurren en una demora de servicio, que no es más que el tiempo requerido para atender una solicitud, es decir, el intervalo de tiempo entre el momento en que un nodo IoT envía una solicitud de servicio y el momento cuando recibe la respuesta para esa solicitud.

La decisión de realizar una tarea se basa en el tiempo de respuesta de un nodo de niebla, el que depende de varios factores: la cantidad de cómputo que se debe realizar en una tarea, el estado de la cola y la capacidad de procesamiento de un nodo de niebla. Se deben tener en cuenta los diferentes tiempos de procesamiento de acuerdo con las diferentes tareas, en otras palabras, se debe distinguir entre las tareas de procesamiento pesado y las tareas de procesamiento ligeras. (Lai, Song, Hwang, & Lai, 2016). Por ejemplo, las solicitudes enviadas por sensores de temperatura a nodos de niebla para calcular la temperatura ambiente promedio, se pueden ver como tareas de procesamiento ligero; del mismo modo, una solicitud de lectura de matrícula de un vehículo, enviado por una cámara de tráfico a los nodos de niebla, es un ejemplo de una tarea de procesamiento pesado.

Cuando los nodos niebla reciben solicitudes de nodos de IoT que participan en una aplicación, deben distinguir entre solicitudes ligeras y pesadas, para ello las solicitudes deben tener en su encabezado un campo que identifique el tipo de solicitud, por ejemplo el campo de clase de tráfico en el encabezado de IPv6.

Existen dos modos de interacción para nodos niebla (Cardellini, Grassi, Lo Presti, & Nardelli, 2015): un modo centralizado, en el que una autoridad central controla la interacción entre los nodos niebla; y un modo distribuido, donde los nodos niebla interactúan con sus nodos vecinos usando un protocolo universal.

El modo centralizado de interacción se puede ver como una orquestación central de recursos, donde está el nodo central conocedor de la topología y el estado de los nodos niebla en un dominio. El modo centralizado es más fácil de implementar, porque no hay necesidad de una comunicación distribuida a través de un protocolo. Además, el nodo central podría ser utilizado como un medio para impulsar las aplicaciones de niebla y realizar las actualizaciones a los nodos de su dominio.

El modo distribuido es más adecuado para escenarios en los que los nodos niebla no son estáticos o cuando la red de niebla se forma de una manera *ad hoc*. Además, en un modo distribuido, no es necesario tener un nodo dedicado para actuar como nodo central, lo que representa una reducción en el costo de implementación, pero una mayor vulnerabilidad ante fallos, al tener un solo punto de fracaso.

Las demoras de propagación se dan a los nodos de niebla como parámetros de entrada. Según el modo de interacción, al recibir una muestra estimada del tiempo de espera, ya sea desde el nodo central o desde otro nodo de niebla, el nodo niebla actualiza el correspondiente tiempo estimado de espera en la tabla de accesibilidad.

La selección del modo distribuido de nodos niebla, como estrategia para proporcionar QoS, es la más eficiente para un escenario tan dinámico que abarca potencialmente miles de millones de dispositivos. La estrategia se basa principalmente en ajustar proporcionalmente la cantidad de controladores dependiendo de la cantidad de recursos (Cardellini et al., 2015).

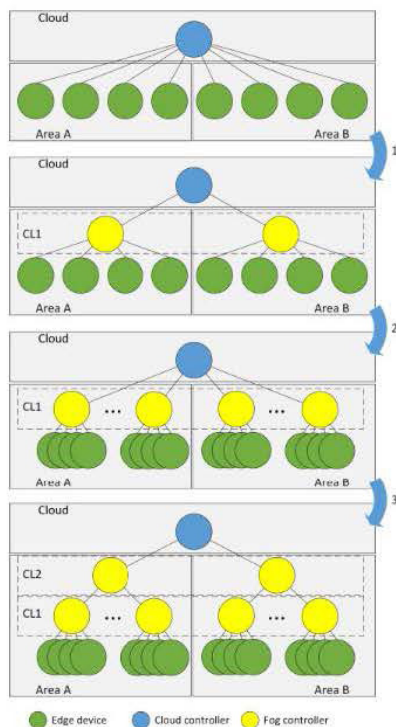


Figure 3. Evolution of the fog nodes / Evolución de los nodos niebla (Chin-Feng, Dong-Yu, & Ren-Hung, 2016)

act as a central one; which represents a reduction in the implementation cost, but a larger failure vulnerability by having a single failing point.

The delays in the propagation are given to the fog nodes as input parameters. As per the interaction mode, when an estimated sample of the waiting time is received –either from the central node or from another one–, the fog node updates the corresponding waiting time in the accessibility table.

The selection of the distributed mode in the fog nodes –as a strategy to provide QoS– is the most efficient for a dynamic scenario entailing billions of connected devices. The strategy is mainly based on adjusting the number of controllers depending on the amount of resources (Cardellini et al., 2015).

In **FIGURE 3**, the blue arrow tagged with number 1 illustrates the deployment of fog controllers when the controller layer is down (CL1 in yellow) to correctly handle the high control delay produced by the fog controller. In addition, when the amount of the border resources grow, the new controllers must be deployed in CL1 to avoid that overwrites in the existing ones, as the arrow tagged with number 2 shows. Nevertheless, the collateral effect is that, when the higher the amount of controllers in CL1 is, the higher the traffic due to the communication between controllers will be, causing a larger average in the delays. To address this, it is possible to implement a new controller layer (CL2) as the arrow tagged with number 3 presents: a new layer with a controller per area to avoid high latency in the communication with the cloud controller. For example, even when the amount of controllers in CL1 is relatively low, the implementation of CL2 –containing 1 single controller per area– can reduce the average communication between areas.

IV. Fog Computing Applications in Smart Cities

The smart and sustainable cities paradigm has gained relevance to face the challenges of the modern urbanization. The ICT use allows higher efficiency in the operations and urban services, improves the life quality of the citizens, and encourages the environmental sustainability.

A smart and sustainable city is an innovative one using the ICT and other means to improve life quality, efficiency in the services and urban operation, and competitiveness. This, by ensuring the satisfaction of needs of the present and future generations relative to economic, social, environmental, and cultural aspects (ITU, 2017).

In the scientific literature, several applications and use cases of fog computing for smart cities are reported. In Bangalore, as part of a pilot project, some surveillance cameras, environmental sensors, edge computing platforms (e.g., Raspberry Pi), and fog computing platforms (e.g., NVIDIA Jetson TX1) have been installed across the city (Varshney & Simmhan, 2017). The use of fog computing for urban monitoring through surveillance cameras is one of the scenarios where better results have been achieved, by reducing the latency when the sharing of processing loads with the cloud is performed. Further, this offers more efficient real-time (or nearly real-time) responses (Gupta et al, 2017). Mohamed, Al-Jaroodi, Jawhar, Lazarova-Molnar, & Mahmoud (2017) describe another fog computing applications in intelligent transportation systems, intelligent energy, intelligent water, and environmental monitoring.

These use cases have presented good practices and learned lessons about the manner to implement fog computing in intelligent cities, and they were analyzed to propose a roadmap seeking to face the implementation of the fog computing within a project starting to be developed in Havana (Cuba). In this project, the experimentation of technological prototypes seeking to transform the city towards a sustainable and intelligent city is proposed.

V. A Roadmap for Mobile Applications in Smart Cities Based on Fog Computing

The Cuban Union of Informatics [UIC, Unión de Informáticos de Cuba] is starting an experimentation project of smart and sustainable cities in Havana, where it should start with the creation of an urban laboratory in old Havana. This project –approved by the National Computation Program of Science, Technology, and Innovation for the 2018-2021 period– includes in its objectives to develop a line of intelligent transportation systems aiming to offer different information types (and recommendations) to the people in a determined zone, establishing a traveler information system using LED displays, tactile screens, or mobile devices.

There are experiences that work as precedent, such as a spatial recommendation system sensible to the traveler context, which recommended sightseeing places through mobile devices with GPS and by using the maps in the Spatial Infrastructure Data of the Cuban Republic (González, Delgado, Capote, & Cruz, 2013) are presen-

En la **FIGURA 3**, la flecha azul 1 ilustra el despliegue de los controladores de niebla al bajar la capa del controlador (CL1 en amarillo) para manejar adecuadamente el alto retardo de control producido por el controlador de la nube. Además, a medida que crece la cantidad de recursos de borde, los nuevos controladores deberán ser desplegados en CL1 para evitar que los existentes se sobrecarguen, como se muestra en la flecha 2. Sin embargo, el efecto colateral es que, cuanto mayor sea la cantidad de controladores en CL1, mayor será el tráfico debido a la comunicación entre los controladores, lo que trae como resultado un mayor promedio de retrasos. Para hacer frente a esto, se puede implementar una nueva capa de controlador (CL2), como indica la flecha 3, una nueva capa con un controlador por área para evitar alta latencia de la comunicación con el controlador de la nube. Por ejemplo, incluso cuando la cantidad de controladores en CL1 es relativamente baja, la implementación de CL2, que contiene un solo controlador por área, puede disminuir la comunicación promedio entre las áreas.

IV. Aplicaciones de *fog computing* en ciudades inteligentes

El paradigma de ciudades inteligentes y sostenibles ha cobrado auge ante los desafíos de la urbanización moderna. El uso de las TIC permite mayor eficacia en las operaciones y en los servicios urbanos, mejora la calidad de vida de los ciudadanos y fomenta la sostenibilidad medioambiental.

Una ciudad inteligente y sostenible es una ciudad innovadora que usa las TIC y otros medios para mejorar la calidad de vida, la eficiencia de los servicios y de la operación urbana, y la competitividad, mientras asegura la satisfacción de las necesidades de las generaciones presentes y futuras con respecto a los aspectos económicos, sociales, ambientales y culturales (ITU, 2017).

En la literatura se reportan múltiples aplicaciones y casos de uso de *fog computing* para ciudades inteligentes. En Bangalore, como parte de un proyecto piloto, se han instalado cámaras de vigilancia, sensores ambientales, plataformas de computación en el borde (e.g., Raspberry Pi) y, en cada cuadra de la ciudad, plataformas de *fog computing* (e.g., NVIDIA Jetson TX1) (Varshney & Simmhan, 2017). El uso de *fog computing* para monitoreo urbano mediante cámaras de vigilancia, es uno de los escenarios donde se observan mejores resultados (Gupta et al, 2017), al disminuir la latencia, compartiendo procesamiento con la nube y ofreciendo respuestas más eficientemente en tiempo real o cercano al real. Mohamed, Al-Jaroodi, Jawhar, Lazarova-Molnar, & Mahmoud, (2017) describen otras aplicaciones de *fog computing* en sistemas inteligentes de transporte, energía inteligente, agua inteligente y monitoreo ambiental.

Estos casos de uso arrojan buenas prácticas y lecciones aprendidas sobre la manera de implementar *fog computing* en ciudades Inteligentes, y fueron analizados para proponer una hoja de ruta con vistas a enfrentar la implementación de *fog computing* en el marco de un proyecto que está comenzando a desarrollarse en La Habana (Cuba) en el que se propone la experimentación de prototipos tecnológicos para hacer la ciudad más inteligente y sostenible.

V. Una hoja de ruta para aplicaciones móviles en ciudades inteligentes basadas en computación en la niebla

La Unión de Informáticos de Cuba está iniciando un proyecto de experimentación de ciudades más inteligentes y sostenibles, que debe comenzar con la creación de un laboratorio urbano en La Habana vieja. Este proyecto, aprobado por el Programa Nacional de Ciencia, Tecnología e Innovación de Informatización para el período 2018-2021, incluye entre sus objetivos desarrollar una línea de sistemas inteligentes de transporte con el objetivo de: ofrecer diferentes tipos de información (y recomendaciones) a las personas de una determinada zona en un sistema de información del viajero, utilizando pantallas de visualización de LED, pantallas táctiles o dispositivos móviles.

Existen experiencias que sirven de antecedente, como un sistema de recomendaciones espacial y sensible al contexto para viajeros (González, Delgado, Capote, & Cruz, 2013) que recomendaba sitios de interés a través de dispositivos móviles con GPS, usando los servicios de mapas en la nube de la Infraestructura de Datos Espaciales de la República de Cuba. Para ello, la señal del GPS y otros datos de un dispositivo móvil de cada vehículo tiene que viajar a un centro de despacho central en la nube y regresar la recomendación al viajero, en un tiempo cercano al real. A pesar de que a nivel de prototipo el sistema funcionó con un rendimiento aceptable, se busca reducir la latencia y aprovechar la configuración distribuida geográficamente de los sistemas de computación en la niebla para su generalización en aplicaciones reales de ciudades inteligentes.

Para construir una nueva versión de sistema de recomendaciones de lugares de interés en la ciudad, con sensibilidad al contexto basado en *fog computing*, se ha conformado una hoja de ruta general, la que incluye los cuatro pasos generales (áreas de investigación) descritos a continuación:

1. Definir una arquitectura de *fog computing* para sistemas de información al viajero en ciudades inteligentes. Como parte de ésta etapa, es necesario descubrir la disponibilidad de recursos capaces de jugar un papel de controlador, configurar posteriormente los posibles recursos descubiertos y, finalmente, crear una topología de controladores al establecer una relación jerárquica entre ellos.
2. Desarrollar los mecanismos necesarios para desplegar los servicios en la capa de la niebla, que se adapten a la aplicación de móviles en cuestión. Este paso incluye explotar estándares existentes y plataformas abiertas (e.g., Docker), extendiéndolas para incluir los elementos necesarios, así como, tecnologías eficientes de virtualización que sean aplicables a la mayor cantidad de nodos físicos posible.
3. Desarrollar servicios de gestión en la niebla, como QoS, QoE [*Quality of Experience*], continuidad del servicio, eficiencia de la red y balance de carga, todo ello en correspondencia con la arquitectura específica definida del sistema de recomendaciones al viajero basado en *fog computing* (paso 1), de forma tal que se encuentre una estrategia óptima para distribuir acciones del servicio de la niebla en relación con los procesamientos en la nube.

ted to the end user. The GPS signal and other data in a mobile device and from the vehicles must be transmitted to a central control point in the cloud and return with the recommendations to the traveler in almost real-time. In spite of the fact that, at the prototype level, the system worked with an acceptable performance, in general terms, the reduction of the latency and to take advantage of the geographically distributed configuration of the fog computing systems in real smart cities applications is the desired scenario.

In order to build a new version of the sightseeing recommendation system in the city and using the fog computing context, a general roadmap has been proposed. This roadmap includes the following four general steps (research tasks):

1. Define a fog computing architecture for information systems to the travelers in smart cities. In this step, it is necessary to discover the resources availability capable to act as controller, to configure the discovered resources, and to create a controller topology by establishing a hierarchical relation amongst them.
2. Develop the necessary mechanisms to deploy the services in the fog layer, which can be adapted to mobile devices. This step also includes to consider existing standards and open platforms (e.g., Docker) and extend them to include the necessary elements, such as efficient virtualization technologies applicable to the majority of physical nodes as possible.
3. Develop fog management systems, such as QoS, QoE [*Quality of Experience*], service continuity, network efficiency, and load balance. All of them according to the defined architecture of the traveler recommendation system using fog computing (step 1). This, pursuing an optimum strategy to distribute fog service actions relative to the cloud processing tasks.
4. Integrate all the components of the recommendation system and test its performance in real scenarios of old Havana.

VI. Conclusions and Future Work

Fog computing complements the cloud paradigm seeking to reduce the latency and the unnecessary internet data traffic using geographically distributed resources closer to the source (sensors). From this new paradigm, a countless number of new applications towards all the society dimensions can emerge.

The new network architectures must use the cloud and fog resources to improve the QoS in the data transmission and reduce the processing delays. For this, the impact of the decisions the fog nodes take relative to the QoS in real time for the NGN [Next Generation Networks] is relevant. This QoS control can be done through the control of the topologies and considering crucial parameters in hierarchical architectures such as the number of layers or the ability to manage each controller.

New challenges will come together with fog computing, however, its low latency and its location at the edge of local networks –with less bandwidth demand–, considering that only what cannot be solved in the fog is sent to the cloud, make it particularly interesting to implement the Internet of Things in an incremental way and learning from the experience. The described roadmap to develop recommendation systems sensible to the context and compatible with mobile environments for smart cities based on fog computing offer a starting point to experiment its advantages and limitations in a real scenario. *ST*

4. Integrar todos los componentes del sistema de recomendaciones de lugares de interés al viajero basado en computación en la niebla y probar su rendimiento en escenarios reales de La Habana vieja.

VI. Conclusiones y trabajo futuro

Fog computing complementa al paradigma de la *cloud*, con el fin de disminuir la latencia y el tráfico innecesario de datos en Internet, mediante el aprovechamiento de recursos distribuidos geográficamente más cerca de la fuente (sensores). A partir de este nuevo paradigma podrá develarse un sinnúmero de nuevas aplicaciones hacia todas las dimensiones de la sociedad.

Las nuevas arquitecturas de red deberán explotar los recursos de la niebla y la nube con el fin de mejorar la QoS en la transmisión de datos y minimizar los retrasos de procesamiento. Para ello, es importante el impacto de las decisiones de los nodos niebla sobre la QoS en tiempo real para las redes de próxima generación [*Next Generation Networks*, NGN], a través del control de las topologías, teniendo en cuenta parámetros cruciales en arquitecturas jerárquicas, como el número de capas o la capacidad de gestión de cada controlador.

Nuevos desafíos vendrán de la mano de la *fog computing*, sin embargo, su baja latencia y su ubicación en el borde de redes locales, con menos demanda de ancho de banda, considerando que solo se envía a la nube lo que no se puede resolver en la niebla, lo hacen particularmente interesante para implementar la Internet de las Cosas de forma incremental e ir aprendiendo de la experiencia. La hoja de ruta descrita para desarrollar sistemas de recomendaciones sensibles al contexto y móviles en ciudades inteligentes basadas en computación en la niebla ofrece un punto de partida para experimentar sus ventajas y limitaciones en un escenario real. *ST*

References / Referencias

- AbuAbdelshkour, M. (2015). *IoT, from cloud to fog computing*. Retrieved from: <https://blogs.cisco.com/perspectives/iot-from-cloud-to-fog-computing>
- Akamai Technologies. (2017). *Q1 2017 State of the Internet / connectivity report*. Retrieved from: <https://www.akamai.com/us/en/multimedia/documents/state-of-the-internet/q1-2017-state-of-the-internet-connectivity-report.pdf>
- Arco Research Group (2017, March 12). *El papel de la tecnología de lógica reconfigurable como respuesta a los retos del Internet de las Cosas*. Retrieved from: <https://arcoresearchgroup.wordpress.com/2017/03/12/el-papel-de-la-tecnologia-de-logica-reconfigurable-como-respuesta-a-los-retos-del-internet-de-las-cosas/>
- Bojanova, I. (2015). *What makes up the Internet of Things?* Retrieved from: <https://www.computer.org/web/sensing-iot/content?g=53926943&type=article&urlTitle=what-are-the-components-of-iot->
- Bonomi, F., Milito, R., Zhu, J., & Addepalli, S. (2012). Fog computing and its role in the Internet of Things. In: *Proceedings of the first edition of the MCC Workshop on Mobile Cloud Computing*, (pp. 13-16). New York, NY: ACM.
- Cardellini, V., Grassi, V., Lo Presti, F., Nardelli, M. (2015). On QoS-aware Scheduling of Data Stream Applications over Fog Computing Infrastructures. In: *2015 IEEE Symposium on Computers and Communication (ISCC)*, (pp. 271-276). doi:10.1109/ISCC.2015.7405527
- Cisco Systems. (2017). *Fog computing and the Internet of things: Extend the cloud to where the things are* [white paper]. Retrieved from: https://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/computing-overview.pdf
- De Fuenmayor, A. (2017, Feb. 20). *Cloud computing, ¿el final de una era?* [blog ThinkingBig]. Retrieved from: <https://aunclic-delastic.blogthinkbig.com/cloud-computing-iot/>
- Gonzalez, G., Delgado, T., Capote, J. L., & Cruz, R. (2013). Context-aware recommender system based on ontologies. In: H. Onsrud, & A. Rajabifard (Eds.), *Spatially enablement in support of economic development and poverty reduction*, (pp. 227-243). Reston, VA: GSDI.
- Gupta, H., Vahid-Dastjerdi, A., Ghosh, S.K., Buyya, R. (2017). FogSim: A toolkit for modeling and simulation of resource management techniques in internet of things, edge and fog computing environments. *Software: Practice and Experience*, 47(9), 1275-1296.
- Hu, P., Dhelim, S., Ning, H. Qiu, T. (2017). Survey on fog computing: Architecture, key technologies, applications and open issues. *Journal of Network and Computer Applications*, 98, 27-42.
- International Telecommunications Union [ITU]. (25/ 4/2017). *The Manizales Manifesto*. Retrieved from: <https://www.itu.int/en/ITU-T/Workshops-and-Seminars/gsw/201704/Documents/Manifesto-Manizales-05-04-2017-Eng-Final.pdf>
- Lai, C. F., Song, D. Y., Hwang, R. H., & Lai, Y. X. (2016). A QoS-aware streaming service over fog computing infrastructures. In *Digital Media Industry & Academic Forum (DMI AF)*, (pp. 94-98). IEEE.
- Mapped: The world according to Internet connection speeds - Telegraph* (2017, April, 9). Retrieved from: <http://www.telegraph.co.uk/travel/maps-and-graphics/countries-with-fastest-internet-connection-speeds/>
- Mohamed, N., Al-Jaroodi, J., Jawhar, I., Lazarova-Molnar, S., & Mahmoud, S. (2017). Smartcityware: A service-oriented middleware for cloud and fog enabled smart city services. *IEEE Acces*, 5, 17576-17588. doi:10.1109/ACCESS.2017.2731382
- Varshney, P., & Simmhan, Y. (2017). Demystifying fog computing: Characterizing architectures, applications and abstractions. In *Fog and Edge Computing (ICFEC), 2017 IEEE 1st International Conference on* (pp. 115-124). IEEE.
- Yousefpour, A., Ishigaki, G., & Jue, J. P. (2017). Fog computing: Towards minimizing delay in the internet of things. In *Edge Computing (EDGE), 2017 IEEE International Conference on* (pp. 17-24). IEEE.

CURRICULUM VITAE

Shouddy Tárano León Informatics Engineer, professor in Computing Sciences and senior consultant in information technologies. Specialized in free and open source technologies for a high scale informatics solutions. His scientific production includes: graphs by computer, information's architecture, smart systems, optimization and high development computing. His working area is focused in informatics systems with a high level of integration and a high level of heterogeneity involving several technology providers including: virtualization, networks, operational systems, security, cloud computing and system's performance analysis / Ingeniero Informático, profesor de Ciencias de la Computación, consultor senior en tecnologías de la información. Especializado en tecnologías libres y de código abierto para soluciones informáticas a gran escala. Su producción científica incluye: gráficas por computadora, arquitectura de la información, sistemas inteligentes, optimización y computación de alto desempeño. Trabaja con sistemas informáticos con elevados niveles de integración y variados grados de heterogeneidad a partir de múltiples proveedores de tecnologías que involucran: virtualización, redes, sistemas de operación, componentes de seguridad, computación en la nube y análisis de rendimiento y desempeño de sistemas.

Tatiana Delgado Fernández Engineer in Automated Systems in Management from former Instituto Politécnico José Antonio Echevarría (Havana, Cuba). She holds a Master's degree in Optimization and Decision Making, and a Ph.D., in Technical Sciences. She is an associated professor at the Business Information Department of the Universidad Tecnológica de La Habana and Vice President of the Unión de Informáticos de Cuba. Her areas of interest are: spatial data infrastructures, Big Data, ontologies, smart cities and IT governance / Graduada de Ingeniería en Sistemas Automatizados en Dirección del entonces Instituto Politécnico José Antonio Echevarría (La Habana, Cuba), ostenta un título de MSc. en Optimización y Toma de Decisiones y el grado de Doctor en Ciencias Técnicas. Es Profesora Titular del Departamento de Informática Empresarial de la Universidad Tecnológica de la Habana (Cuba) y Vicepresidenta de la Unión de Informáticos de Cuba. Sus áreas de interés son: infraestructura de datos espaciales, Big Data, ontologías, smart cities y gobierno de TI.

Alejandro Luar Pérez Colomé Student of fifth year of Telecommunications and Electronics Engineering at the Universidad Tecnológica de la Habana (Cuba), where he has participated in several scientific working days. His work about the new kinds of telecommunications attacks obtained an award in one of them. He is currently a member of the Telematics Research Group of the University. His interests in research include the Internet of Things, fog computing, Big Data and content distribution networks / Estudiante de quinto año de Ingeniería en Telecomunicaciones y Electrónica en la Universidad Tecnológica de La Habana (Cuba). Ha participado en varias jornadas científicas de la Universidad, donde su trabajo sobre los nuevos tipos de ataques a las redes de telecomunicaciones obtuvo fue premiado. Actualmente es miembro del Grupo de Investigación en Telemática de la Universidad. Sus intereses en investigación incluyen Internet de las Cosas, computación en la niebla, Big Data y redes de distribución de contenidos.